NCKU CSIE 2026 Capstone Project Exhibition

UAV Autonomous Navigation Based on Monocular Depth Estimation and Hybrid Path Planning

Advisor: Prof. Yeim-Kuan Chang

Students: Cheng-yu Pu · Liang-wei Chen

KEY WORDS : Monocular Vision · Depth Anything V2 · APF-Informed-RRT · Dynamic Window Approach · AirSim*

Overview

This project was Built Upon Open-Source ROS Noetic Framework:

Monocular-SLAM-Drone

Gaigalas Jonas, Perkauskas Linas, Gricius Henrikas, Kanapickas Tomas & Kriščiūnas Andrius

"A Framework for Autonomous UAV Navigation Based on Monocular Depth Estimation" Drones, MDPI, 2025

drones

MDPI

Article

A Framework for Autonomous UAV Navigation Based on Monocular Depth Estimation

Jonas Gaigalas *, Linas Perkauskas, Henrikas Gricius, Tomas Kanapickas and Andrius Kriščiūnas

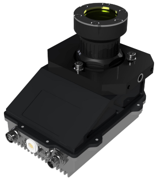
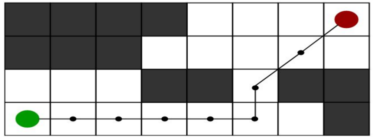
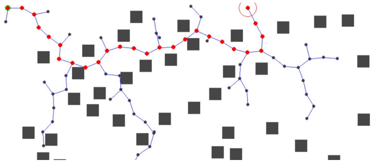
Department of Applied Informatics, Faculty of Informatics, Kaunas University of Technology, Studentu St. 50, LT-51368 Kaunas, Lithuania
* Correspondence: jonas.gaigalas@ktu.edu

Our Key Contributions

We extend the framework by replacing the original A* path planner with an APF-guided Informed-RRT* algorithm, integrated with a fine-tuned monocular depth model.

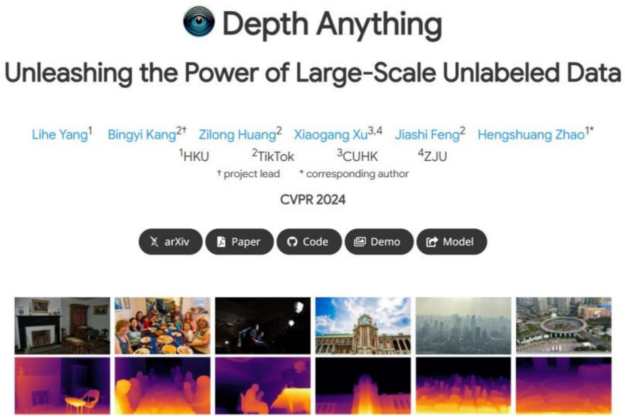
The system is validated across 720 automated flights in Microsoft AirSim. You can find our work in Github: https://github.com/pukyle/Depth-Vision-UAV-Autopilot

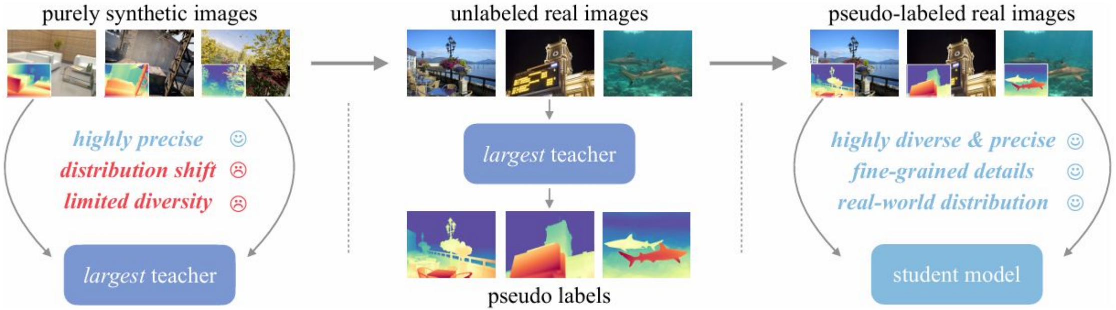
Research Motivation

	Current Bottlenecks	Proposed Solutions
Hardware	 **LiDAR:** Provides high accuracy but is too heavy and expensive for small UAVs.	 **RGB Camera & Monocular Vision:** Lightweight and cost-effective for small UAVs.
Software	 **A* Pathfinding Algorithm:** A* is rigid in 3D dynamic spaces.	 **Hybrid Planning:** Combines APF-Informed-RRT* for global path efficiency.

Objective: Combine monocular depth estimation with hybrid path planning for lightweight, autonomous UAV navigation.

Depth Estimation Foundation: Depth Anything V2

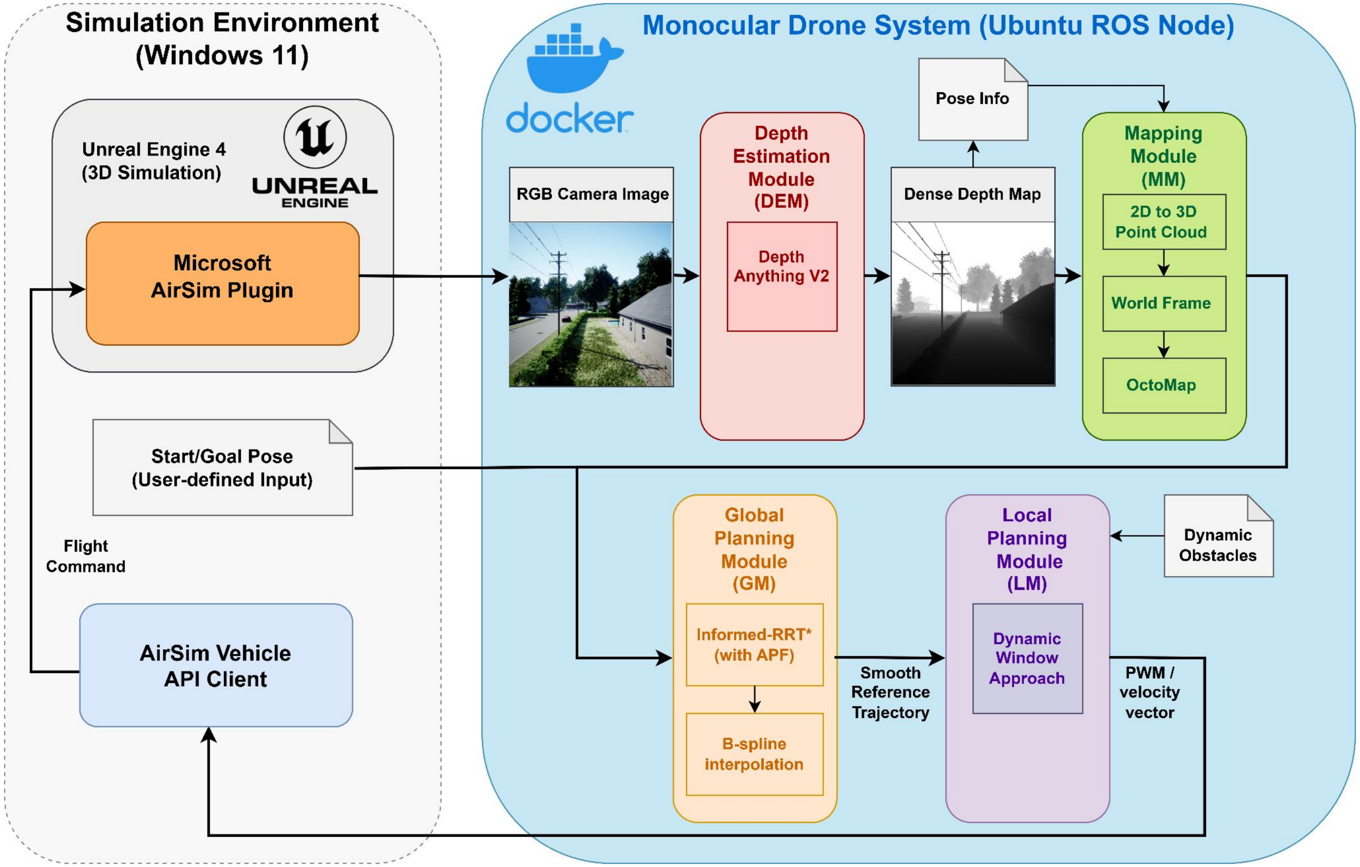




Depth Anything V2 is a **transformer-based monocular depth estimation model** trained on a large-scale dataset comprising 1.5 million labeled and 62 million unlabeled images.

- ➢ Over **10× faster** and more efficient than diffusion-based depth models,
- ➢ Provided in multiple sizes (from 25M to 1.3B parameters) to support various application scales,
- ➢ Fine-tuned to produce highly accurate metric depth estimates, and
- ➢ Accompanied by a new benchmark dataset with diverse scenes

System Architecture

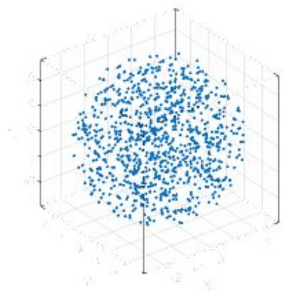


Perception & 3D Mapping Pipeline

Depth Estimation

Depth Anything V2 convert standard RGB input into dense depth maps.

Transformation

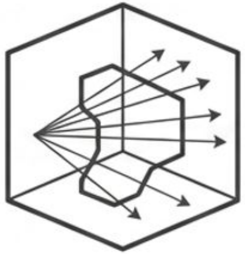


2D depth pixels are mathematically projected into a structured 3D Point Cloud

Occupancy Grid

OctoMap builds a dynamic, probabilistic 3D collision map from the point cloud

Optimization



Empty space is actively identified and cleared using 3D ray-casting

Constructing the 3D Navigable Grid

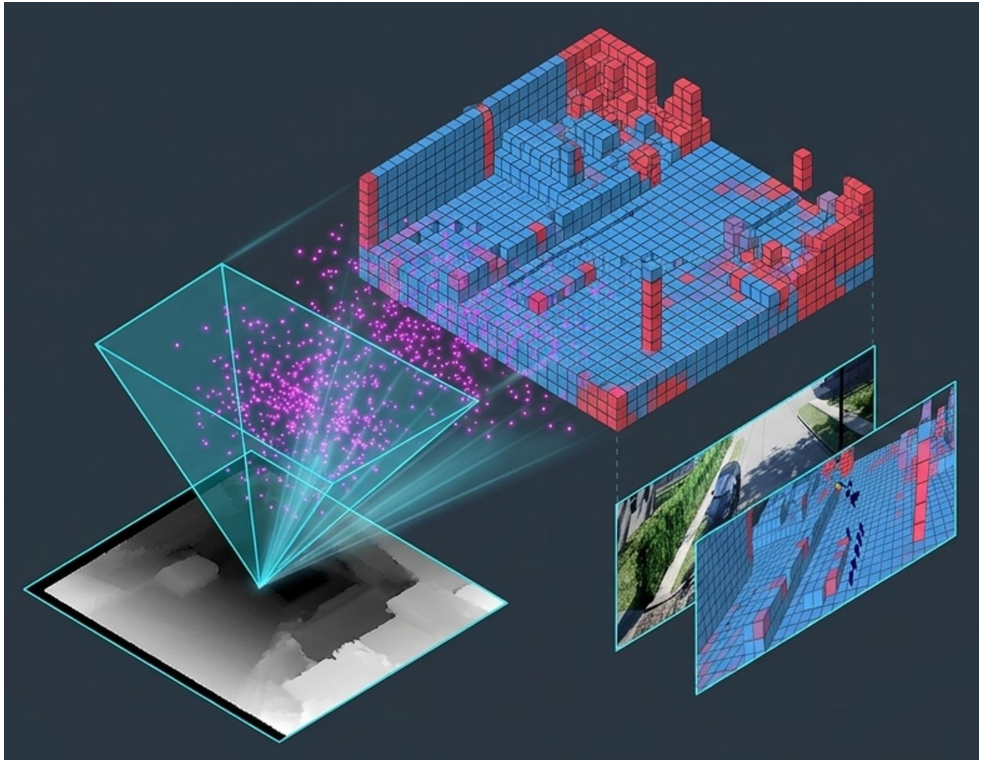


Fig. 3D Mapping Pipeline Visualization

Depth map → Point cloud

The grayscale depth map (bottom-left) is back-projected through the camera frustum into a 3D magenta point cloud.

Red voxels — Occupied

High-probability occupied cells detected from the point cloud. It must treat these as solid obstacles.

Blue voxels — Free space

Cells confirmed empty via 3D ray-casting. The global planner searches exclusively through this navigable volume.

Synthetic Dataset Generation for Monocular Vision

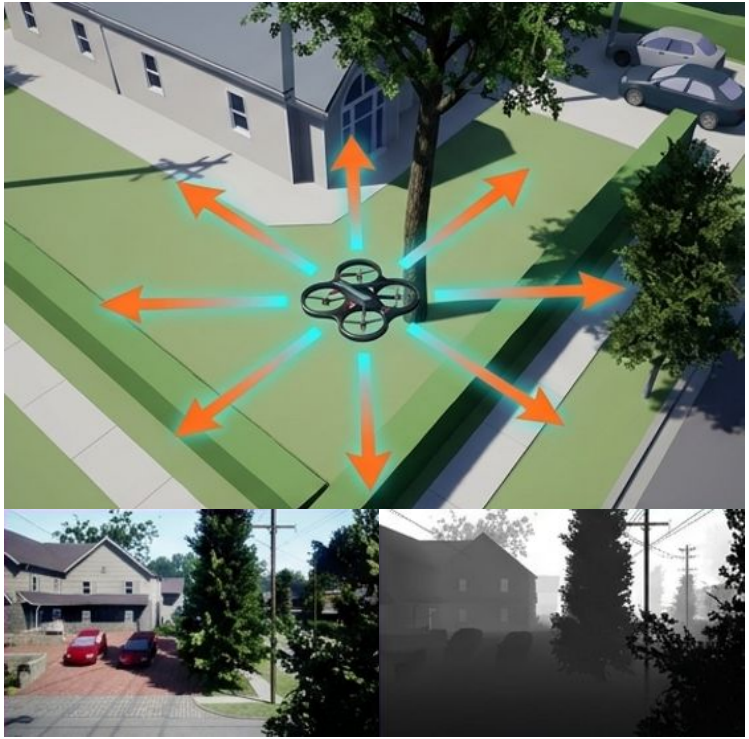


Fig. AirSim multi-height data capture

Scale Drift Problem

Pre-trained monocular models suffer from inherent scale-drift and lighting sensitivity.

Dataset Scale

Generated 50,000 synthetic RGB-Depth pairs across varied AirSim simulation environments.

Collection Method

Captured autonomously at multiple heights (1m–9m) across 8 distinct camera angles.

Geometric Diversity

Provided the model with vast geometrical variations to eliminate scale ambiguity.

Overcoming Scale Drift: Depth Anything V2 Fine-Tuning

Raw RGB Frame

Pre-Trained Depth (Scale Drift)

Fine-Tuned Depth

Architecture

Adapted the Vision Transformer (ViT-S) encoder specifically for structural scene analysis.

BerHu Loss

Employed BerHu loss to handle large depth discrepancies, combining L1 & L2 terms to prevent gradient explosions.

Outcome

Model infers statistically accurate metric depths without requiring explicit scale correction.

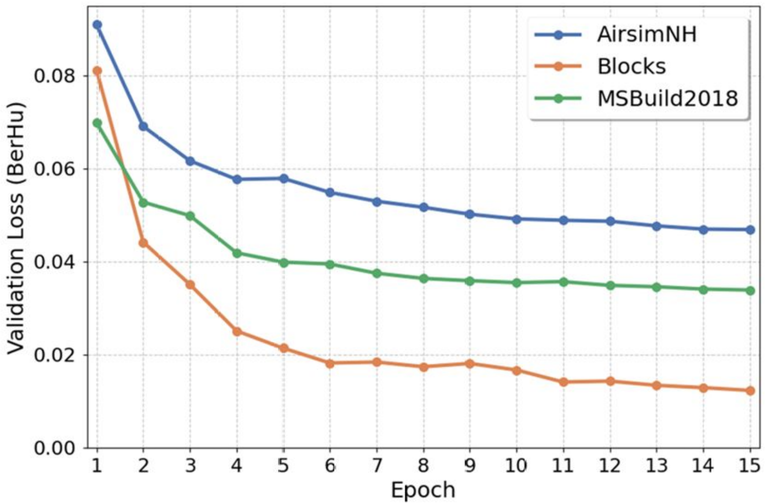
Empirical Validation: Depth Estimation Accuracy

Mean Absolute Percentage Error — MAPE (%)

Environment	Pre-Trained Model	Fine-Tuned Model
Blocks	70.09%	8.04%
AirSim Neighborhood	120.45%	33.41%
MSBuild2018	121.94%	32.86%

Fine-tuning yielded a dramatic drop in depth estimation errors across all simulation scenarios.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|A_i - F_i|}{A_i}$$



Global Planning: Why APF-Informed-RRT*?

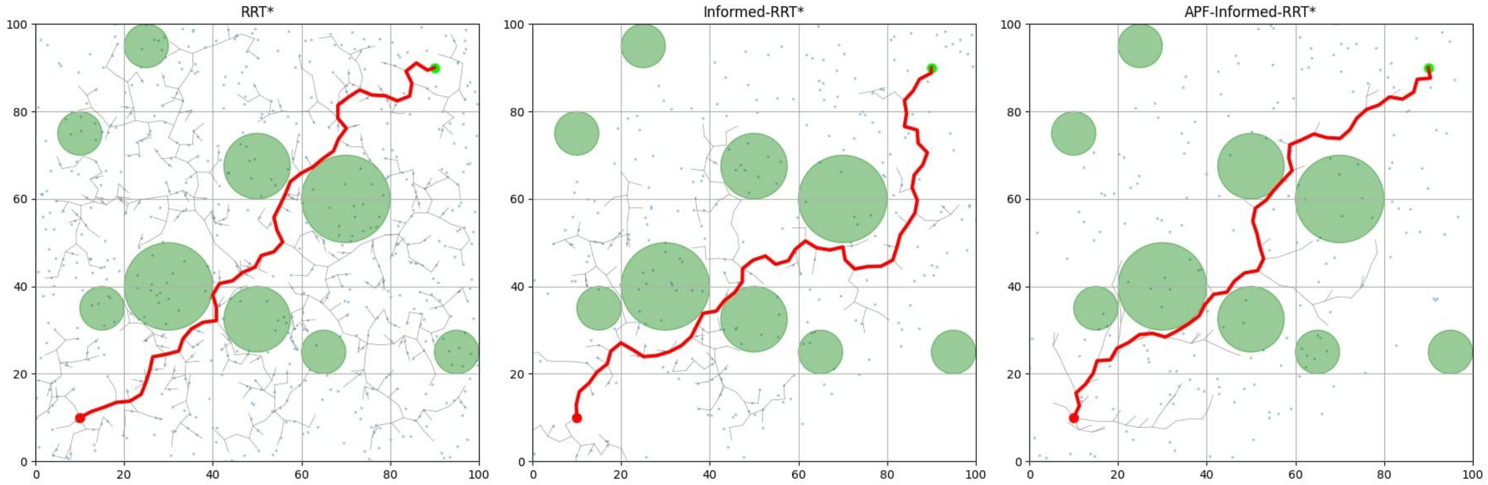


Fig. Comparison of search space: RRT ⟶ Informed-RRT* ⟶ APF-Informed-RRT**

Informed-RRT*

Constraints random sampling to a heuristic elliptical region, drastically reducing wasted exploration.

Artificial Potential Fields

Attractive force pulls nodes toward the goal; repulsive force steers sampling away from obstacles.

Key Insight

APF guidance reduces redundant nodes and accelerates convergence. B-spline interpolation then smooths the raw path.

Comparison of Three RRT* Variants

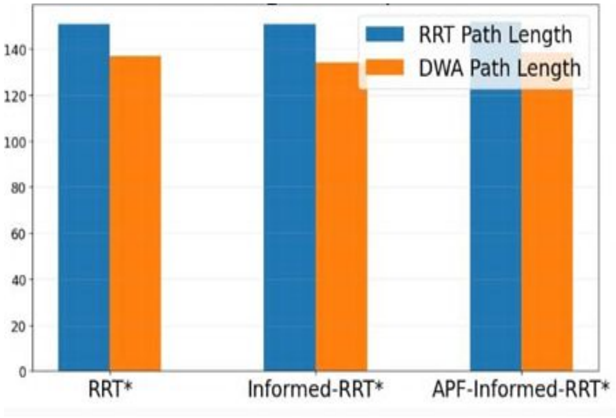


Fig.Path Length Comparison (m)

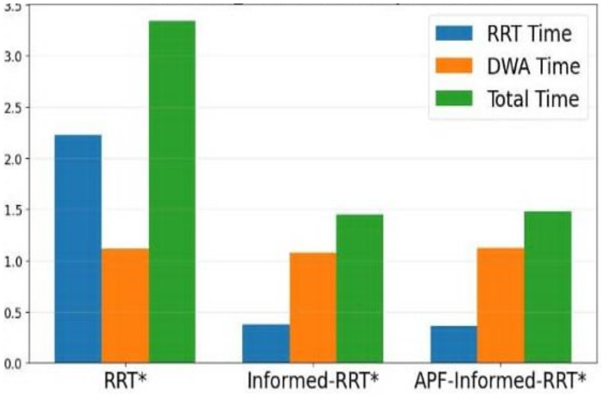


Fig.Planning Time Comparison (s)

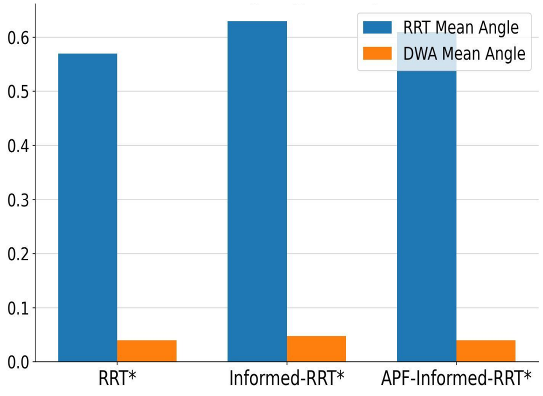


Fig.Mean Turning Angle (rad)

	RRT*	Informed-RRT*	APF-Informed-RRT*
Search Space	**Full random exploration**	**Elliptical heuristic region**	**Ellipse + APF obstacle bias**
Convergence	**Slow, many wasted nodes**	**Faster, fewer samples needed**	**Fastest, nodes are concentrated**
Path Quality	**Long, erratic**	**Shorter, cleaner**	**Shortest**

Local Planning: B-Spline Smoothing & DWA

Kinematic Smoothing — B-Spline

- Smooths raw RRT* waypoints into kinematically feasible curves.
- Removes sharp turns, ensuring the UAV can physically execute the path.
- Extracts key nodes as input targets for the DWA local planner.

Dynamic Window Approach (DWA)

- Real-time obstacle avoidance via safe velocity pair (v, ω) evaluation.
- Instantly adapts to both unmapped static and dynamic obstacles.

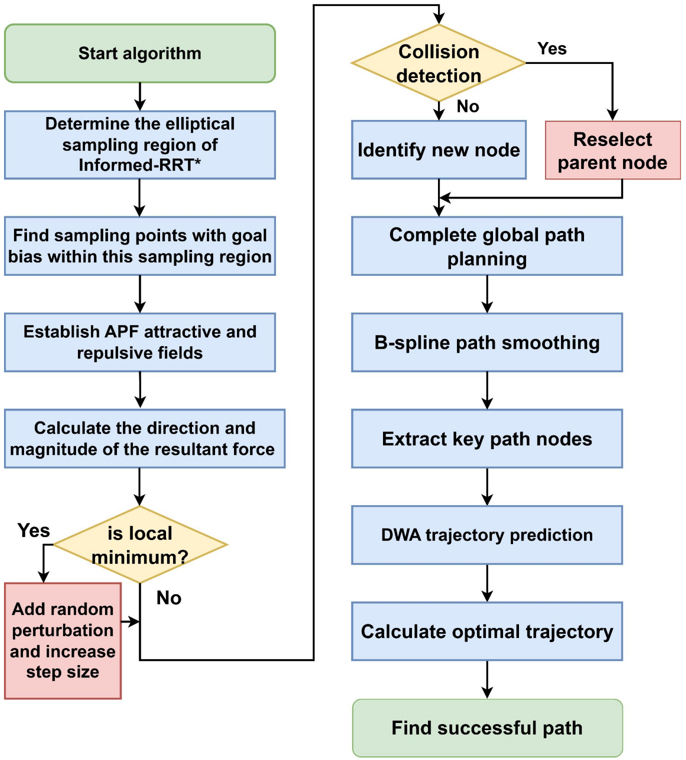
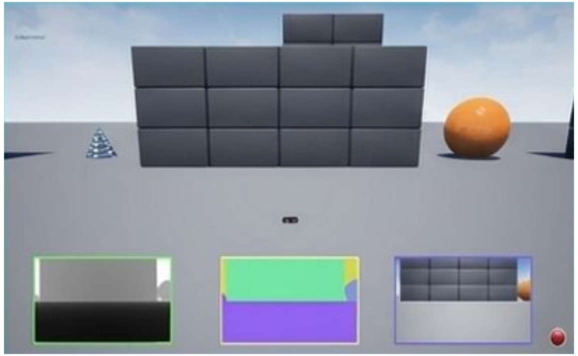


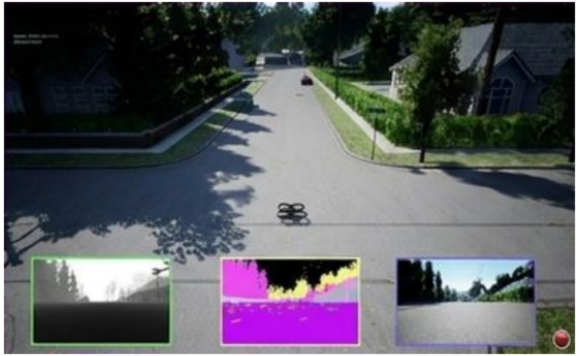
Fig. Hybrid Global-Local Planning Algorithm Flow Chart

Simulation Environments in Microsoft AirSim



Blocks

Baseline testing environment for static obstacle avoidance and physical kinematic limits.



AirSim Neighborhood

Introduces dynamic elements, fences, and complex backyard geometries.

MSBuild 2018

Extreme complexity environment — tests algorithm robustness against dense foliage and narrow gaps. (Ongoing)

Experimental Setup & Procedure

This project conducted a total of **720 autonomous flights**, spanning:

2 simulation environments (Blocks, AirSim NH) × **2** depth sources (Ground-truth depth, Estimated depth) × **2** planning methods (A*, APF-Informed-RRT*) × **3** routes × **3** voxel sizes (0.5 / 1.0 / 2.0 m) × **10** repeated trials.

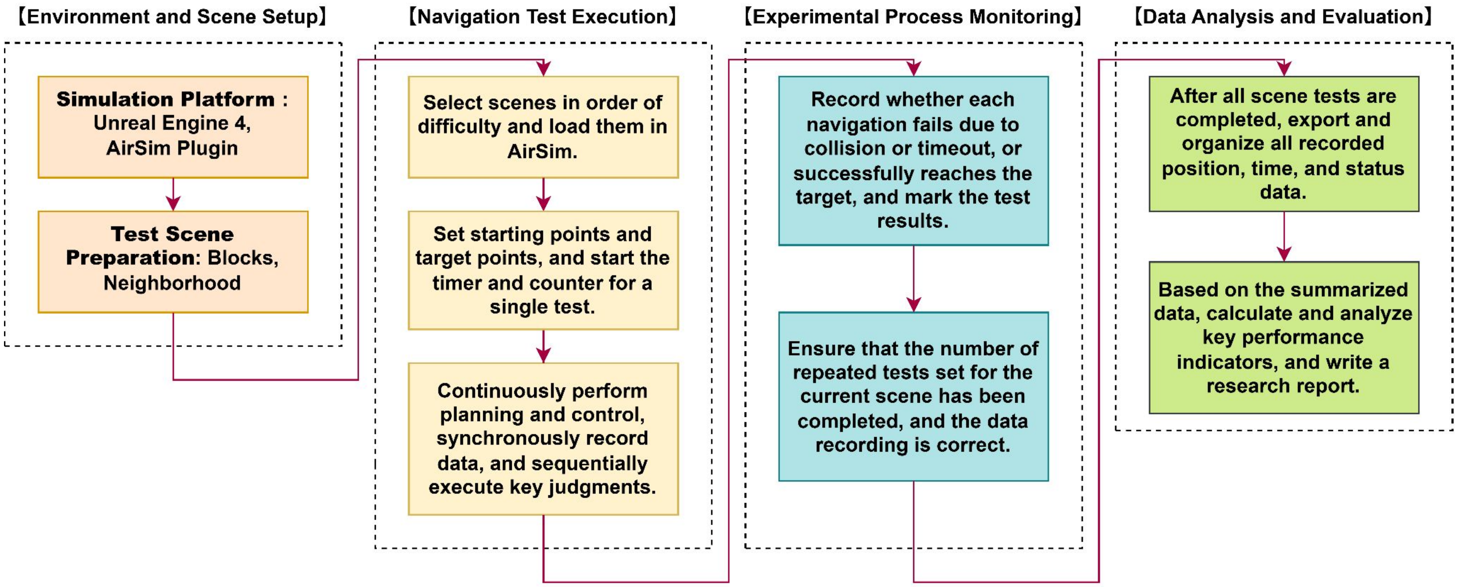
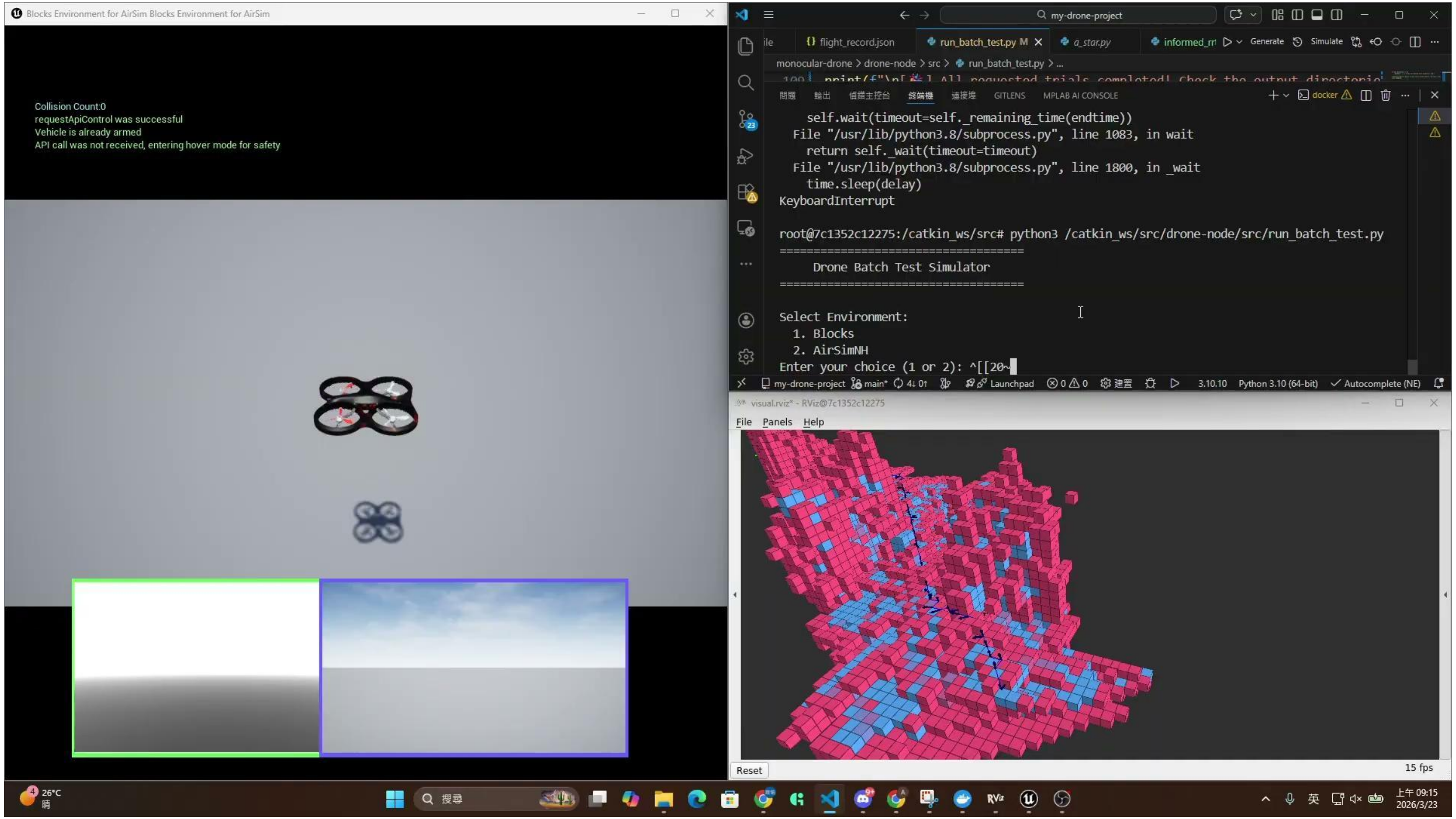


Fig. Experimental Flow Chart

Experimental Demo

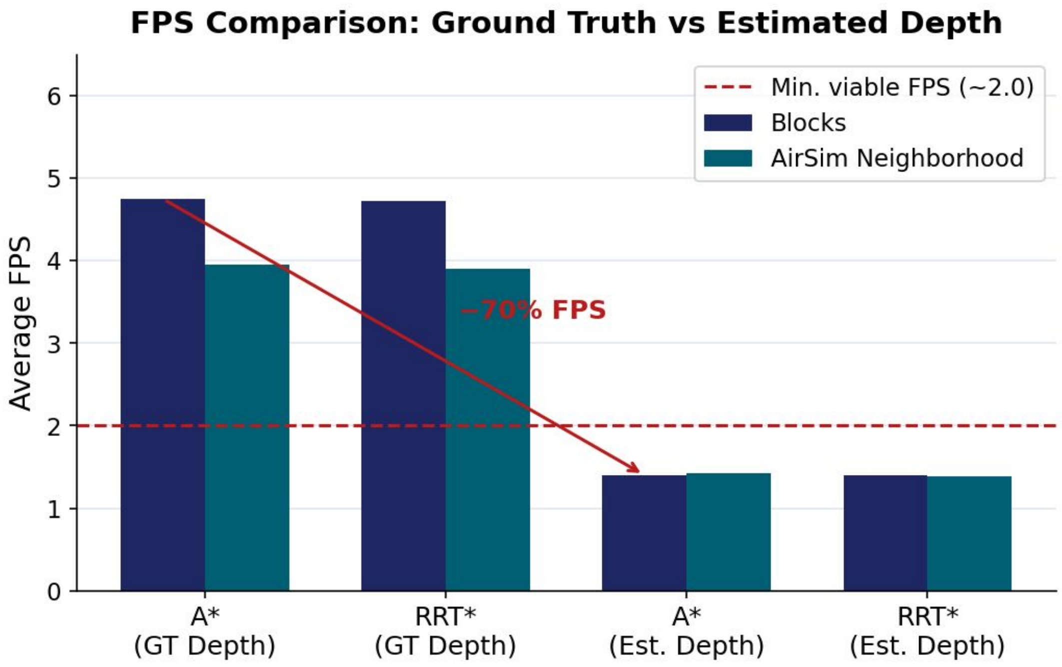


Experimental Results : APF-informed-RRT* vs A*

Method	Depth Source	Blocks Success	AirSim NH Success	Path Ratio	Jerk/f	FPS
A*	Ground Truth	69/90 (76.7%)	36/90 (40.0%)	1.117	0.922	4.75
RRT*	Ground Truth	**73/90 (81.1%)**	**42/90 (46.7%)**	1.115	0.973	4.73
A*	Estimated	40/90 (44.4%)	**0/90 (0.0%)**	1.356	1.770	1.40
RRT*	Estimated	**42/90 (46.7%)**	**0/90 (0.0%)**	1.341	1.808	1.40

Table. Overall Navigation Results Across Environments and Depth Sources

Experimental Results : The Estimated Depth Challenge



Root Cause: Depth Anything V2 inference **consumes ~70% of compute budget** → map update rate drops below safe threshold.

FPS drops 70 %

Neural depth inference drops system FPS from **~4.7 to ~1.4**, causing stale occupancy maps and failed re planning.

Zero Successes in AirSim NH

Both A* and RRT* fail entirely in the complex environment — too slow to react to obstacles.

Partial in Blocks

Simpler geometry allows 44–47% success despite low FPS, but path quality degrades significantly.

Conclusion & Future Work

Core Implementation

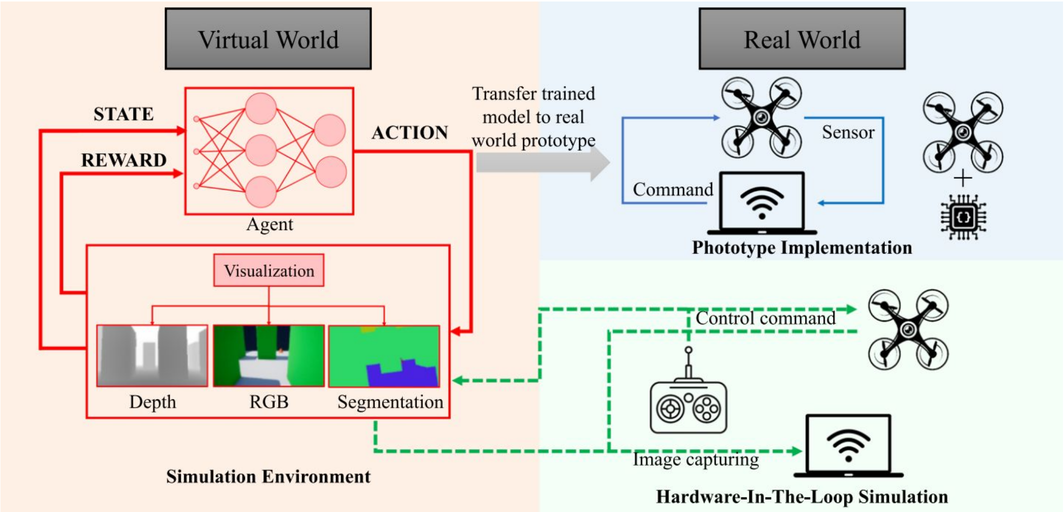
✓ Demonstrated a robust, vision-only hybrid navigation framework in complex 3D simulations.

✓ APF-Informed-RRT* outperforms baseline A* in both simple and complex 3D environments.

Next Steps

Hardware Deployment: Port the edge-computing architecture to NVIDIA Jetson Orin for onboard real-time inference.

Sim-to-Real Transfer: Validate the full system in physical indoor and outdoor UAV field environments.



References

1. Arampatzakis, V., et al. (2024). Monocular Depth Estimation: A Thorough Review. IEEE Trans. Pattern Anal. Mach. Intell., 46(4), 2396–2414. https://doi.org/10.1109/TPAMI.2023.3330944
2. Yang, L., et al. (2024). Depth Anything V2. NeurIPS 2024. https://arxiv.org/abs/2406.09414
3. Gaigalas, J., et al. (2025). A Framework for Autonomous UAV Navigation Based on Monocular Depth Estimation. Drones, 9(4), 236. https://doi.org/10.3390/drones9040236
4. LaValle, S. M. (1998). Rapidly-exploring random trees: A new tool for path planning. Technical Report 98-11, Iowa State University.
5. Karaman, S., & Frazzoli, E. (2011). Sampling-based Algorithms for Optimal Motion Planning. Int. J. Robot. Res., 30(7), 846–894. https://doi.org/10.1177/0278364911406761
6. Gammell, J. D., et al. (2014). Informed RRT*: Optimal Sampling-based Path Planning. IEEE/RSJ IROS 2014, 2997–3004. https://doi.org/10.1109/IROS.2014.6942976
7. Wu, T., et al. (2024). A Dynamic Path Planning Method for UAVs Based on Improved Informed-RRT* Fused Dynamic Windows. Drones, 8(10), 539. https://doi.org/10.3390/drones8100539
8. Fox, D., Burgard, W., & Thrun, S. (1997). The dynamic window approach to collision avoidance. IEEE Robot. Autom. Mag., 4(1), 23–33. https://doi.org/10.1109/100.580977
9. Shah, S., et al. (2017). AirSim: High-Fidelity Visual and Physical Simulation for Autonomous Vehicles. Field and Service Robotics. https://arxiv.org/abs/1705.05065